

# WO9914737

Publication Title:

## METHOD FOR DETECTING SPEECH ACTIVITY

Abstract:

### Abstract of WO9914737

The invention concerns a method whereby the digital speech signal ( $s$ ) processed by successive frames is subjected to noise suppression taking into account noise estimations included in the signal, updated for each frame based on at least one degree of speech activity (gamma  $n,i$ ). The method consists in carrying out an a priori noise suppression of each frame speech signal on the basis of the noise estimations obtained while processing at least one previous frame, and analysing the energy variations of the signal which has been subjected to an a priori noise suppression to detect the degree of speech activity of said frame.

Data supplied from the esp@cenet database - Worldwide b59

---

Courtesy of <http://v3.espacenet.com>



DEMANDE INTERNATIONALE PUBLIEE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

***UNIQUEMENT A TITRE D'INFORMATION***

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publient des demandes internationales en vertu du PCT.

AL	Albanie	ES	Espagne	LS	Lesotho	SI	Slovénie
AM	Arménie	FI	Finlande	LT	Lituanie	SK	Slovaquie
AT	Autriche	FR	France	LU	Luxembourg	SN	Sénégal
AU	Australie	GA	Gabon	LV	Lettonie	SZ	Swaziland
AZ	Azerbaïjan	GB	Royaume-Uni	MC	Monaco	TD	Tchad
BA	Bosnie-Herzégovine	GE	Géorgie	MD	République de Moldova	TG	Togo
BB	Barbade	GH	Ghana	MG	Madagascar	TJ	Tadjikistan
BE	Belgique	GN	Guinée	MK	Ex-République yougoslave de Macédoine	TM	Turkménistan
BF	Burkina Faso	GR	Grèce	ML	Mali	TR	Turquie
BG	Bulgarie	HU	Hongrie	MN	Mongolie	TT	Trinité-et-Tobago
BJ	Bénin	IE	Irlande	MR	Mauritanie	UA	Ukraine
BR	Brésil	IL	Israël	MW	Malawi	UG	Ouganda
BY	Bélarus	IS	Islande	MX	Mexique	US	Etats-Unis d'Amérique
CA	Canada	IT	Italie	NE	Niger	UZ	Ouzbékistan
CF	République centrafricaine	JP	Japon	NL	Pays-Bas	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norvège	YU	Yougoslavie
CH	Suisse	KG	Kirghizistan	NZ	Nouvelle-Zélande	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	République populaire démocratique de Corée	PL	Pologne		
CM	Cameroun	KR	République de Corée	PT	Portugal		
CN	Chine	KZ	Kazakhstan	RO	Roumanie		
CU	Cuba	LC	Sainte-Lucie	RU	Fédération de Russie		
CZ	République tchèque	LI	Liechtenstein	SD	Soudan		
DE	Allemagne	LK	Sri Lanka	SE	Suède		
DK	Danemark	LR	Libéria	SG	Singapour		
EE	Estonie						

PROCEDE DE DETECTION D'ACTIVITE VOCALE

La présente invention concerne les techniques numériques de traitement de signaux de parole. Elle concerne plus particulièrement les techniques faisant 5 appel à une détection d'activité vocale afin d'effectuer des traitements différenciés selon que le signal supporte ou non une activité vocale.

Les techniques numériques en question relèvent de domaines variés : codage de la parole pour la transmission 10 ou le stockage, reconnaissance de la parole, diminution du bruit, annulation d'écho...

Les méthodes de détection d'activité vocale ont pour principale difficulté la distinction entre l'activité vocale et le bruit qui l'accompagne. Le recours à une 15 technique de débruitage classique ne permet pas de traiter cette difficulté, puisque ces techniques font elles-mêmes appel à des estimations du bruit qui dépendent du degré d'activité vocale du signal.

Un but principal de la présente invention est 20 d'améliorer la robustesse au bruit des méthodes de détection d'activité vocale.

L'invention propose ainsi un procédé de détection d'activité vocale dans un signal de parole numérique traité par trames successives, dans lequel on soumet le 25 signal de parole à un débruitage en tenant compte d'estimations du bruit compris dans le signal, mises à jour pour chaque trame d'une manière dépendante d'au moins un degré d'activité vocale déterminé pour ladite trame. Selon l'invention, on procède à un débruitage a priori du 30 signal de parole de chaque trame sur la base d'estimations du bruit obtenues lors du traitement d'au moins une trame précédente, et on analyse les variations d'énergie du signal débruité a priori pour détecter le degré d'activité vocale de ladite trame.

35 Le fait de procéder à la détection d'activité vocale (selon une méthode qui peut généralement être toute méthode connue) sur la base d'un signal débruité a priori

- 2 -

améliore sensiblement les performances de cette détection lorsque le bruit environnant est relativement important.

Dans la suite de la présente description, on illustrera le procédé de détection d'activité vocale selon 5 l'invention dans un système de débruitage d'un signal de parole. On comprendra que ce procédé peut trouver des applications dans de nombreux autres types de traitement numérique de la parole dans lesquels on souhaite disposer d'une information sur le degré d'activité vocale du signal 10 traité : codage, reconnaissance, annulation d'écho...

D'autres particularités et avantages de la présente invention apparaîtront dans la description ci-après d'exemples de réalisation non limitatifs, en référence aux dessins annexés, dans lesquels :

15 - la figure 1 est un schéma synoptique d'un système de débruitage mettant en œuvre la présente invention ;

20 - les figures 2 et 3 sont des organigrammes de procédures utilisées par un détecteur d'activité vocale du système de la figure 1 ;

- la figure 4 est un diagramme représentant les états d'un automate de détection d'activité vocale ;

- la figure 5 est un graphique illustrant les variations d'un degré d'activité vocale ;

25 - la figure 6 est un schéma synoptique d'un module de surestimation du bruit du système de la figure 1 ;

- la figure 7 est un graphique illustrant le calcul d'une courbe de masquage ; et

30 - la figure 8 est un graphique illustrant l'exploitation des courbes de masquage dans le système de la figure 1.

Le système de débruitage représenté sur la figure 1 traite un signal numérique de parole s. Un module de fenêtrage 10 met ce signal s sous forme de fenêtres ou 35 trames successives, constituées chacune d'un nombre N d'échantillons de signal numérique. De façon classique, ces trames peuvent présenter des recouvrements mutuels. Dans la suite de la présente description, on considérera,

- 3 -

sans que ceci soit limitatif, que les trames sont constituées de  $N=256$  échantillons à une fréquence d'échantillonnage  $F_e$  de 8 kHz, avec une pondération de Hamming dans chaque fenêtre, et des recouvrements de 50% entre fenêtres consécutives.

La trame de signal est transformée dans le domaine fréquentiel par un module 11 appliquant un algorithme classique de transformée de Fourier rapide (TFR) pour calculer le module du spectre du signal. Le module 11 délivre alors un ensemble de  $N=256$  composantes fréquentielles du signal de parole, notées  $S_{n,f}$ , où n désigne le numéro de la trame courante, et f une fréquence du spectre discret. Du fait des propriétés des signaux numériques dans le domaine fréquentiel, seuls les  $N/2=128$  premiers échantillons sont utilisés.

Pour calculer les estimations du bruit contenu dans le signal s, on n'utilise pas la résolution fréquentielle disponible en sortie de la transformée de Fourier rapide, mais une résolution plus faible, déterminée par un nombre I de bandes de fréquences couvrant la bande  $[0, F_e/2]$  du signal. Chaque bande i ( $1 \leq i \leq I$ ) s'étend entre une fréquence inférieure  $f(i-1)$  et une fréquence supérieure  $f(i)$ , avec  $f(0)=0$ , et  $f(I)=F_e/2$ . Ce découpage en bandes de fréquences peut être uniforme ( $f(i)-f(i-1)=F_e/2I$ ). Il peut également être non uniforme (par exemple selon une échelle de barks). Un module 12 calcule les moyennes respectives des composantes spectrales  $S_{n,f}$  du signal de parole par bandes, par exemple par une pondération uniforme telle que :

$$30 \quad S_{n,i} = \frac{1}{f(i) - f(i-1)} \sum_{f \in [f(i-1), f(i)]} S_{n,f} \quad (1)$$

Ce moyennage diminue les fluctuations entre les bandes en moyennant les contributions du bruit dans ces bandes, ce qui diminuera la variance de l'estimateur de bruit. En outre, ce moyennage permet une forte diminution de la complexité du système.

- 4 -

Les composantes spectrales moyennées  $s_{n,i}$  sont adressées à un module 15 de détection d'activité vocale et à un module 16 d'estimation du bruit. Ces deux modules 15, 16 fonctionnent conjointement, en ce sens que des degrés 5 d'activité vocale  $\gamma_{n,i}$  mesurés pour les différentes bandes par le module 15 sont utilisés par le module 16 pour estimer l'énergie à long terme du bruit dans les différentes bandes, tandis que ces estimations à long terme  $\hat{B}_{n,i}$  sont utilisées par le module 15 pour procéder à 10 un débruitage a priori du signal de parole dans les différentes bandes pour déterminer les degrés d'activité vocale  $\gamma_{n,i}$ .

Le fonctionnement des modules 15 et 16 peut correspondre aux organigrammes représentés sur les figures 15 2 et 3.

Aux étapes 17 à 20, le module 15 procède au débruitage a priori du signal de parole dans les différentes bandes  $i$  pour la trame de signal  $n$ . Ce débruitage a priori est effectué selon un processus 20 classique de soustraction spectrale non linéaire à partir d'estimations du bruit obtenues lors d'une ou plusieurs trames précédentes. A l'étape 17, le module 15 calcule, avec la résolution des bandes  $i$ , la réponse en fréquence  $H_{p,n,i}$  du filtre de débruitage a priori, selon la formule :

$$25 \quad H_{p,n,i} = \frac{s_{n,i} - \alpha'_{n-\tau_1,i} \cdot \hat{B}_{n-\tau_1,i}}{s_{n-\tau_2,i}} \quad (2)$$

où  $\tau_1$  et  $\tau_2$  sont des retards exprimés en nombre de trames ( $\tau_1 \geq 1$ ,  $\tau_2 \geq 0$ ), et  $\alpha'_{n,i}$  est un coefficient de surestimation du bruit dont la détermination sera expliquée plus loin. Le retard  $\tau_1$  peut être fixe (par exemple  $\tau_1=1$ ) ou variable. 30 Il est d'autant plus faible qu'on est confiant dans la détection d'activité vocale.

Aux étapes 18 à 20, les composantes spectrales

- 5 -

$\hat{E}p_{n,i}$  sont calculées selon :

$$\hat{E}p_{n,i} = \max\left\{Hp_{n,i} \cdot S_{n,i}, \beta p_i \cdot \hat{B}_{n-\tau_1,i}\right\} \quad (3)$$

où  $\beta p_i$  est un coefficient de plancher proche de 0, servant classiquement à éviter que le spectre du signal débruité 5 prenne des valeurs négatives ou trop faibles qui provoqueraient un bruit musical.

Les étapes 17 à 20 consistent donc essentiellement à soustraire du spectre du signal une estimation, majorée par le coefficient  $\alpha'_{n-\tau_1,i}$ , du spectre du bruit estimé a 10 priori.

A l'étape 21, le module 15 calcule l'énergie du signal débruité a priori dans les différentes bandes i pour la trame n :  $E_{n,i} = \hat{E}p_{n,i}^2$ . Il calcule aussi une moyenne globale  $E_{n,0}$  de l'énergie du signal débruité a 15 priori, par une somme des énergies par bande  $E_{n,i}$ , pondérée par les largeurs de ces bandes. Dans les notations ci-dessous, l'indice  $i=0$  sera utilisé pour désigner la bande globale du signal.

Aux étapes 22 et 23, le module 15 calcule, pour 20 chaque bande i ( $0 \leq i \leq I$ ), une grandeur  $\Delta E_{n,i}$  représentant la variation à court terme de l'énergie du signal débruité dans la bande i, ainsi qu'une valeur à long terme  $\bar{E}_{n,i}$  de l'énergie du signal débruité dans la bande i. La grandeur  $\Delta E_{n,i}$  peut être calculée par une formule simplifiée de

25 dérivation :  $\Delta E_{n,i} = \left| \frac{E_{n-4,i} + E_{n-3,i} - E_{n-1,i} - E_{n,i}}{10} \right|$ . Quant à l'énergie à long terme  $\bar{E}_{n,i}$ , elle peut être calculée à l'aide d'un facteur d'oubli  $B1$  tel que  $0 < B1 < 1$ , à savoir  $\bar{E}_{n,i} = B1 \cdot \bar{E}_{n-1,i} + (1-B1) \cdot E_{n,i}$ .

- 6 -

Après avoir calculé les énergies  $E_{n,i}$  du signal débruité, ses variations à court terme  $\Delta E_{n,i}$  et ses valeurs à long terme  $\bar{E}_{n,i}$  de la manière indiquée sur la figure 2, le module 15 calcule, pour chaque bande  $i$  ( $0 \leq i \leq I$ ), une valeur  $p_i$  représentative de l'évolution de l'énergie du signal débruité. Ce calcul est effectué aux étapes 25 à 36 de la figure 3, exécutées pour chaque bande  $i$  entre  $i=0$  et  $i=I$ . Ce calcul fait appel à un estimateur à long terme de l'enveloppe du bruit  $b_{ai}$ , à un estimateur interne  $b_i$  et à un compteur de trames bruitées  $b_i$ .

A l'étape 25, la grandeur  $\Delta E_{n,i}$  est comparée à un seuil  $\epsilon_1$ . Si le seuil  $\epsilon_1$  n'est pas atteint, le compteur  $b_i$  est incrémenté d'une unité à l'étape 26. A l'étape 27, l'estimateur à long terme  $b_{ai}$  est comparé à la valeur de l'énergie lissée  $\bar{E}_{n,i}$ . Si  $b_{ai} \geq \bar{E}_{n,i}$ , l'estimateur  $b_{ai}$  est pris égal à la valeur lissée  $\bar{E}_{n,i}$  à l'étape 28, et le compteur  $b_i$  est remis à zéro. La grandeur  $p_i$ , qui est prise égale au rapport  $b_{ai}/\bar{E}_{n,i}$  (étape 36), est alors égale à 1.

Si l'étape 27 montre que  $b_{ai} < \bar{E}_{n,i}$ , le compteur  $b_i$  est comparé à une valeur limite  $b_{max}$  à l'étape 29. Si  $b_i > b_{max}$ , le signal est considéré comme trop stationnaire pour supporter de l'activité vocale. L'étape 28 précitée, qui revient à considérer que la trame ne comporte que du bruit, est alors exécutée. Si  $b_i \leq b_{max}$  à l'étape 29, l'estimateur interne  $b_i$  est calculé à l'étape 33 selon :

$$b_i = (1-Bm) \cdot \bar{E}_{n,i} + Bm \cdot b_{ai} \quad (4)$$

Dans cette formule,  $Bm$  représente un coefficient de mise à jour compris entre 0,90 et 1. Sa valeur diffère selon l'état d'un automate de détection d'activité vocale

- 7 -

(étapes 30 à 32). Cet état  $\delta_{n-1}$  est celui déterminé lors du traitement de la trame précédente. Si l'automate est dans un état de détection de parole ( $\delta_{n-1}=2$  à l'étape 30), le coefficient  $B_m$  prend une valeur  $B_{mp}$  très proche de 1 pour que l'estimateur du bruit soit très faiblement mis à jour en présence de parole. Dans le cas contraire, le coefficient  $B_m$  prend une valeur  $B_{ms}$  plus faible, pour permettre une mise à jour plus significative de l'estimateur de bruit en phase de silence. A l'étape 34, l'écart  $b_{ai}-b_i$  entre l'estimateur à long terme et l'estimateur interne du bruit est comparé à un seuil  $\epsilon_2$ . Si le seuil  $\epsilon_2$  n'est pas atteint, l'estimateur à long terme  $b_{ai}$  est mis à jour avec la valeur de l'estimateur interne  $b_i$  à l'étape 35. Sinon, l'estimateur à long terme  $b_{ai}$  reste inchangé. On évite ainsi que de brutales variations dues à un signal de parole conduisent à une mise à jour de l'estimateur de bruit.

Après avoir obtenu les grandeurs  $\rho_i$ , le module 15 procède aux décisions d'activité vocale à l'étape 37. Le module 15 met d'abord à jour l'état de l'automate de détection selon la grandeur  $\rho_0$  calculée pour l'ensemble de la bande du signal. Le nouvel état  $\delta_n$  de l'automate dépend de l'état précédent  $\delta_{n-1}$  et de  $\rho_0$ , de la manière représentée sur la figure 4.

Quatre états sont possibles :  $\delta=0$  détecte le silence, ou absence de parole ;  $\delta=2$  détecte la présence d'une activité vocale ; et les états  $\delta=1$  et  $\delta=3$  sont des états intermédiaires de montée et de descente. Lorsque l'automate est dans l'état de silence ( $\delta_{n-1}=0$ ), il y reste si  $\rho_0$  ne dépasse pas un premier seuil  $SE_1$ , et il passe dans l'état de montée dans le cas contraire. Dans l'état de montée ( $\delta_{n-1}=1$ ), il revient dans l'état de silence si

- 8 -

$\rho_0$  est plus petit que le seuil SE1, il passe dans l'état de parole si  $\rho_0$  est plus grand qu'un second seuil SE2 plus grand que le seuil SE1, et il reste dans l'état de montée si  $SE1 \leq \rho_0 \leq SE2$ . Lorsque l'automate est dans l'état de parole ( $\delta_{n-1}=2$ ), il y reste si  $\rho_0$  dépasse un troisième seuil SE3 plus petit que le seuil SE2, et il passe dans l'état de descente dans le cas contraire. Dans l'état de descente ( $\delta_{n-1}=3$ ), l'automate revient dans l'état de parole si  $\rho_0$  est plus grand que le seuil SE2, il revient 10 dans l'état de silence si  $\rho_0$  est en deçà d'un quatrième seuil SE4 plus petit que le seuil SE2, et il reste dans l'état de descente si  $SE4 \leq \rho_0 \leq SE2$ .

A l'étape 37, le module 15 calcule également les degrés d'activité vocale  $\gamma_{n,i}$  dans chaque bande  $i \geq 1$ . Ce 15 degré  $\gamma_{n,i}$  est de préférence un paramètre non binaire, c'est-à-dire que la fonction  $\gamma_{n,i} = g(\rho_i)$  est une fonction variant continûment entre 0 et 1 en fonction des valeurs prises par la grandeur  $\rho_i$ . Cette fonction a par exemple l'allure représentée sur la figure 5.

Le module 16 calcule les estimations du bruit par bande, qui seront utilisées dans le processus de débruitage, en utilisant les valeurs successives des composantes  $S_{n,i}$  et des degrés d'activité vocale  $\gamma_{n,i}$ . Ceci correspond aux étapes 40 à 42 de la figure 3. A 25 l'étape 40, on détermine si l'automate de détection d'activité vocale vient de passer de l'état de montée à l'état de parole. Dans l'affirmative, les deux dernières estimations  $\hat{B}_{n-1,i}$  et  $\hat{B}_{n-2,i}$  précédemment calculées pour chaque bande  $i \geq 1$  sont corrigées conformément à la valeur 30 de l'estimation précédente  $\hat{B}_{n-3,i}$ . Cette correction est effectuée pour tenir compte du fait que, dans la phase de

- 9 -

montée ( $\delta=1$ ), les estimations à long terme de l'énergie du bruit dans le processus de détection d'activité vocale (étapes 30 à 33) ont pu être calculées comme si le signal ne comportait que du bruit ( $B_m=B_{ms}$ ), de sorte qu'elles 5 risquent d'être entachées d'erreur.

A l'étape 42, le module 16 met à jour les estimations du bruit par bande selon les formules :

$$\tilde{B}_{n,i} = \lambda_B \cdot \hat{B}_{n-1,i} + (1-\lambda_B) \cdot S_{n,i} \quad (5)$$

$$\hat{B}_{n,i} = \gamma_{n,i} \cdot \hat{B}_{n-1,i} + (1-\gamma_{n,i}) \cdot \tilde{B}_{n,i} \quad (6)$$

10 où  $\lambda_B$  désigne un facteur d'oubli tel que  $0 < \lambda_B < 1$ . La formule (6) met en évidence la prise en compte du degré d'activité vocale non binaire  $\gamma_{n,i}$ .

15 Comme indiqué précédemment, les estimations à long terme du bruit  $\hat{B}_{n,i}$  font l'objet d'une surestimation, par un module 45 (figure 1), avant de procéder au débruitage par soustraction spectrale non linéaire. Le module 45 calcule le coefficient de surestimation  $\alpha'_{n,i}$  précédemment évoqué, ainsi qu'une estimation majorée  $\hat{B}'_{n,i}$  qui correspond essentiellement à  $\alpha'_{n,i} \cdot \hat{B}_{n,i}$ .

20 L'organisation du module de surestimation 45 est représentée sur la figure 6. L'estimation majorée  $\hat{B}'_{n,i}$  est obtenue en combinant l'estimation à long terme  $\hat{B}_{n,i}$  et une mesure  $\Delta B_{n,i}^{\max}$  de la variabilité de la composante du bruit dans la bande  $i$  autour de son estimation à long terme. 25 Dans l'exemple considéré, cette combinaison est, pour l'essentiel, une simple somme réalisée par un additionneur 46. Ce pourrait également être une somme pondérée.

Le coefficient de surestimation  $\alpha'_{n,i}$  est égal au rapport entre la somme  $\hat{B}_{n,i} + \Delta B_{n,i}^{\max}$  délivrée par

- 10 -

l'additionneur 46 et l'estimation à long terme retardée  $\hat{B}_{n-\tau_3,i}$  (diviseur 47), plafonné à une valeur limite  $\alpha_{\max}$ , par exemple  $\alpha_{\max}=4$  (bloc 48). Le retard  $\tau_3$  sert à corriger le cas échéant, dans les phases de montée ( $\delta=1$ ), la valeur 5 du coefficient de surestimation  $\alpha'_{n,i}$ , avant que les estimations à long terme aient été corrigées par les étapes 40 et 41 de la figure 3 (par exemple  $\tau_3=3$ ).

L'estimation majorée  $\hat{B}'_{n,i}$  est finalement prise égale à  $\alpha'_{n,i} \cdot \hat{B}_{n-\tau_3,i}$  (multiplieur 49).

10 La mesure  $\Delta B_{n,i}^{\max}$  de la variabilité du bruit reflète la variance de l'estimateur de bruit. Elle est obtenue en fonction des valeurs de  $s_{n,i}$  et de  $\hat{B}_{n,i}$  calculées pour un certain nombre de trames précédentes sur lesquelles le signal de parole ne présente pas d'activité vocale dans la 15 bande i. C'est une fonction des écarts  $|s_{n-k,i} - \hat{B}_{n-k,i}|$  calculés pour un nombre K de trames de silence ( $n-k \leq n$ ). Dans l'exemple représenté, cette fonction est simplement le maximum (bloc 50). Pour chaque trame n, le degré d'activité vocale  $\gamma_{n,i}$  est comparé à un seuil (bloc 51) 20 pour décider si l'écart  $|s_{n,i} - \hat{B}_{n,i}|$ , calculé en 52-53, doit ou non être chargé dans une file d'attente 54 de K emplacements organisée en mode premier entré-premier sorti (FIFO). Si  $\gamma_{n,i}$  ne dépasse pas le seuil (qui peut être égal à 0 si la fonction g() a la forme de la figure 5), la 25 FIFO 54 n'est pas alimentée, tandis qu'elle l'est dans le cas contraire. La valeur maximale contenue dans la FIFO 54 est alors fournie comme mesure de variabilité  $\Delta B_{n,i}^{\max}$ .

La mesure de variabilité  $\Delta B_{n,i}^{\max}$  peut, en variante, être obtenue en fonction des valeurs  $s_{n,f}$  (et non  $s_{n,i}$ ) et

- 11 -

$\hat{B}_{n,i}$ . On procède alors de la même manière, sauf que la FIFO 54 contient non pas  $|S_{n-k,i} - \hat{B}_{n-k,i}|$  pour chacune des bandes i, mais plutôt  $\max_{f \in [f(i-1), f(i)]} |S_{n-k,f} - \hat{B}_{n-k,i}|$ .

Grâce aux estimations indépendantes des 5 fluctuations à long terme du bruit  $\hat{B}_{n,i}$  et de sa variabilité à court terme  $\Delta B_{n,i}^{\max}$ , l'estimateur majoré  $\hat{B}'_{n,i}$  procure une excellente robustesse aux bruits musicaux du procédé de débruitage.

Une première phase de la soustraction spectrale 10 est réalisée par le module 55 représenté sur la figure 1. Cette phase fournit, avec la résolution des bandes i ( $1 \leq i \leq I$ ), la réponse en fréquence  $H_{n,i}^1$  d'un premier filtre de débruitage, en fonction des composantes  $S_{n,i}$  et  $\hat{B}_{n,i}$  et des coefficients de surestimation  $\alpha'_{n,i}$ . Ce calcul peut 15 être effectué pour chaque bande i selon la formule :

$$H_{n,i}^1 = \frac{\max\{S_{n,i} - \alpha'_{n,i} \cdot \hat{B}_{n,i}, \beta_i^1 \cdot \hat{B}_{n,i}\}}{S_{n-\tau4,i}} \quad (7)$$

où  $\tau4$  est un retard entier déterminé tel que  $\tau4 \geq 0$  (par exemple  $\tau4=0$ ). Dans l'expression (7), le coefficient  $\beta_i^1$  représente, comme le coefficient  $\beta p_i$  de la formule (3), un 20 plancher servant classiquement à éviter les valeurs négatives ou trop faibles du signal débruité.

De façon connue (EP-A-0 534 837), le coefficient de surestimation  $\alpha'_{n,i}$  pourrait être remplacé dans la formule (7) par un autre coefficient égal à une fonction 25 de  $\alpha'_{n,i}$  et d'une estimation du rapport signal-sur-bruit (par exemple  $S_{n,i}/\hat{B}_{n,i}$ ), cette fonction étant décroissante

- 12 -

selon la valeur estimée du rapport signal-sur-bruit. Cette fonction est alors égale à  $\alpha'_{n,i}$  pour les valeurs les plus faibles du rapport signal-sur-bruit. En effet, lorsque le signal est très bruité, il n'est a priori pas utile de diminuer le facteur de surestimation. Avantageusement, cette fonction décroît vers zéro pour les valeurs les plus élevées du rapport signal/bruit. Ceci permet de protéger les zones les plus énergétiques du spectre, où le signal de parole est le plus significatif, la quantité soustraite du signal tendant alors vers zéro.

Cette stratégie peut être affinée en l'appliquant de manière sélective aux harmoniques de la fréquence tonale (« pitch ») du signal de parole lorsque celui-ci présente une activité vocale.

Ainsi, dans la réalisation représentée sur la figure 1, une seconde phase de débruitage est réalisée par un module 56 de protection des harmoniques. Ce module calcule, avec la résolution de la transformée de Fourier, la réponse en fréquence  $H_{n,f}^2$  d'un second filtre de débruitage en fonction des paramètres  $H_{n,i}^1$ ,  $\alpha'_{n,i}$ ,  $\hat{B}_{n,i}$ ,  $\delta_n$ ,  $S_{n,i}$  et de la fréquence tonale  $f_p = F_e/T_p$  calculée en dehors des phases de silence par un module d'analyse harmonique 57. En phase de silence ( $\delta_n = 0$ ), le module 56 n'est pas en service, c'est-à-dire que  $H_{n,f}^2 = H_{n,i}^1$  pour chaque fréquence  $f$  d'une bande  $i$ . Le module 57 peut appliquer toute méthode connue d'analyse du signal de parole de la trame pour déterminer la période  $T_p$ , exprimée comme un nombre entier ou fractionnaire d'échantillons, par exemple une méthode de prédiction linéaire.

La protection apportée par le module 56 peut consister à effectuer, pour chaque fréquence  $f$  appartenant à une bande  $i$  :

- 13 -

$$\begin{cases} H_{n,f}^2 = 1 & \text{si} \\ & \left\{ S_{n,i} - \alpha'_{n,i} \cdot \hat{B}_{n,i} > \beta_i^2 \cdot \hat{B}_{n,i} \right. \\ H_{n,f}^2 = H_{n,f}^1 & \text{sinon} \\ & \left. \text{et } \exists \eta \text{ entier } / |f - \eta \cdot f_p| \leq \Delta f / 2 \right\} \end{cases} \quad (8)$$

$$(9)$$

$\Delta f = F_e / N$  représente la résolution spectrale de la transformée de Fourier. Lorsque  $H_{n,f}^2 = 1$ , la quantité soustraite de la composante  $S_{n,f}$  sera nulle. Dans ce  
5 calcul, les coefficients de plancher  $\beta_i^2$  (par exemple  $\beta_i^2 = \beta_i^1$ ) expriment le fait que certaines harmoniques de la fréquence tonale  $f_p$  peuvent être masquées par du bruit, de sorte qu'il n'est pas utile de les protéger.

Cette stratégie de protection est de préférence  
10 appliquée pour chacune des fréquences les plus proches des harmoniques de  $f_p$ , c'est-à-dire pour  $\eta$  entier quelconque.

Si on désigne par  $\delta f_p$  la résolution fréquentielle avec laquelle le module d'analyse 57 produit la fréquence tonale estimée  $f_p$ , c'est-à-dire que la fréquence tonale  
15 réelle est comprise entre  $f_p - \delta f_p / 2$  et  $f_p + \delta f_p / 2$ , alors l'écart entre la  $\eta$ -ième和谐 de la fréquence tonale réelle est son estimation  $\eta \times f_p$  (condition (9)) peut aller jusqu'à  $\pm \eta \times \delta f_p / 2$ . Pour les valeurs élevées de  $\eta$ , cet écart peut être supérieur à la demi-résolution spectrale  $\Delta f / 2$  de  
20 la transformée de Fourier. Pour tenir compte de cette incertitude et garantir la bonne protection des harmoniques de la fréquence tonale réelle, on peut protéger chacune des fréquences de l'intervalle  
[ $\eta \times f_p - \eta \times \delta f_p / 2$ ,  $\eta \times f_p + \eta \times \delta f_p / 2$ ], c'est-à-dire remplacer la  
25 condition (9) ci-dessus par :

$$\exists \eta \text{ entier } / |f - \eta \cdot f_p| \leq (\eta \cdot \delta f_p + \Delta f) / 2 \quad (9')$$

- 14 -

Cette façon de procéder (condition (9')) présente un intérêt particulier lorsque les valeurs de  $\eta$  peuvent être grandes, notamment dans le cas où le procédé est utilisé dans un système à bande élargie.

5 Pour chaque fréquence protégée, la réponse en fréquence corrigée  $H_{n,f}^2$  peut être égale à 1 comme indiqué ci-dessus, ce qui correspond à la soustraction d'une quantité nulle dans le cadre de la soustraction spectrale, c'est-à-dire à une protection complète de la fréquence en question. Plus généralement, cette réponse en fréquence 10 corrigée  $H_{n,f}^2$  pourrait être prise égale à une valeur comprise entre 1 et  $H_{n,f}^1$  selon le degré de protection souhaité, ce qui correspond à la soustraction d'une quantité inférieure à celle qui serait soustraite si la 15 fréquence en question n'était pas protégée.

Les composantes spectrales  $S_{n,f}^2$  d'un signal débruité sont calculées par un multiplicateur 58 :

$$S_{n,f}^2 = H_{n,f}^2 \cdot S_{n,f} \quad (10)$$

20 Ce signal  $S_{n,f}^2$  est fourni à un module 60 qui calcule, pour chaque trame  $n$ , une courbe de masquage en appliquant un modèle psychoacoustique de perception auditive par l'oreille humaine.

25 Le phénomène de masquage est un principe connu du fonctionnement de l'oreille humaine. Lorsque deux fréquences sont entendues simultanément, il est possible que l'une des deux ne soit plus audible. On dit alors qu'elle est masquée.

Il existe différentes méthodes pour calculer des 30 courbes de masquage. On peut par exemple utiliser celle développée par J.D. Johnston («Transform Coding of Audio Signals Using Perceptual Noise Criteria », IEEE Journal on Selected Area in Communications, Vol. 6, No. 2, février 1988). Dans cette méthode, on travaille dans l'échelle fréquentielle des barks. La courbe de masquage

- 15 -

est vue comme la convolution de la fonction d'étalement spectral de la membrane basilaire dans le domaine bark avec le signal excitateur, constitué dans la présente application par le signal  $S_{n,f}^2$ . La fonction d'étalement spectral peut être modélisée de la manière représentée sur la figure 7. Pour chaque bande de bark, on calcule la contribution des bandes inférieures et supérieures convoluées par la fonction d'étalement de la membrane basilaire :

$$10 \quad C_{n,q} = \sum_{q'=0}^{q-1} \frac{S_{n,q'}^2}{(10^{10/10})^{(q-q')}} + \sum_{q'=q+1}^Q \frac{S_{n,q'}^2}{(10^{25/10})^{(q'-q)}} \quad (11)$$

où les indices  $q$  et  $q'$  désignent les bandes de bark ( $0 \leq q, q' \leq Q$ ), et  $S_{n,q'}^2$  représente la moyenne des composantes  $S_{n,f}^2$  du signal excitateur débruité pour les fréquences discrètes  $f$  appartenant à la bande de bark  $q'$ .

15        Le seuil de masquage  $M_{n,q}$  est obtenu par le module 60 pour chaque bande de bark  $q$ , selon la formule :

$$M_{n,q} = C_{n,q}/R_q \quad (12)$$

où  $R_q$  dépend du caractère plus ou moins voisé du signal. De façon connue, une forme possible de  $R_q$  est :

$$20 \quad 10.\log_{10}(R_q) = (A+q).\chi + B.(1-\chi) \quad (13)$$

avec  $A=14,5$  et  $B=5,5$ .  $\chi$  désigne un degré de voisement du signal de parole, variant entre zéro (pas de voisement) et 1 (signal fortement voisé). Le paramètre  $\chi$  peut être de la forme connue :

$$25 \quad \chi = \min\left\{\frac{SFM}{SF_{max}}, 1\right\} \quad (12)$$

où SFM représente, en décibels, le rapport entre la moyenne arithmétique et la moyenne géométrique de l'énergie des bandes de bark, et  $SF_{max}=-60$  dB.

Le système de débruitage comporte encore un module 30 62 qui corrige la réponse en fréquence du filtre de

- 16 -

débruitage, en fonction de la courbe de masquage  $M_{n,q}$  calculée par le module 60 et des estimations majorées  $\hat{B}'_{n,i}$  calculées par le module 45. Le module 62 décide du niveau de débruitage qui doit réellement être atteint.

5 En comparant l'enveloppe de l'estimation majorée du bruit avec l'enveloppe formée par les seuils de masquage  $M_{n,q'}$ , on décide de ne débruiter le signal que dans la mesure où l'estimation majorée  $\hat{B}'_{n,i}$  dépasse la courbe de masquage. Ceci évite de supprimer inutilement du  
10 bruit masqué par de la parole.

La nouvelle réponse  $H_{n,f}^3$ , pour une fréquence  $f$  appartenant à la bande  $i$  définie par le module 12 et à la bande de bark  $q$ , dépend ainsi de l'écart relatif entre l'estimation majorée  $\hat{B}'_{n,i}$  de la composante spectrale  
15 correspondante du bruit et la courbe de masquage  $M_{n,q'}$ , de la manière suivante :

$$H_{n,f}^3 = 1 - \left(1 - H_{n,f}^2\right) \cdot \max\left\{\frac{\hat{B}'_{n,i} - M_{n,q'}}{\hat{B}'_{n,i}}, 0\right\} \quad (14)$$

En d'autres termes, la quantité soustraite d'une composante spectrale  $S_{n,f'}$  dans le processus de  
20 soustraction spectrale ayant la réponse fréquentielle  $H_{n,f}^3$ , est sensiblement égale au minimum entre d'une part la quantité soustraite de cette composante spectrale dans le processus de soustraction spectrale ayant la réponse fréquentielle  $H_{n,f}^2$ , et d'autre part la fraction de  
25 l'estimation majorée  $\hat{B}'_{n,i}$  de la composante spectrale correspondante du bruit qui, le cas échéant, dépasse la courbe de masquage  $M_{n,q'}$ .

La figure 8 illustre le principe de la correction appliquée par le module 62. Elle montre schématiquement un

- 17 -

exemple de courbe de masquage  $M_{n,q}$  calculée sur la base des composantes spectrales  $S_{n,f}^2$  du signal débruité, ainsi que l'estimation majorée  $\hat{B}'_{n,i}$  du spectre du bruit. La quantité finalement soustraite des composantes  $S_{n,f}$  sera 5 celle représentée par les zones hachurées, c'est-à-dire limitée à la fraction de l'estimation majorée  $\hat{B}'_{n,i}$  des composantes spectrales du bruit qui dépasse la courbe de masquage.

Cette soustraction est effectuée en multipliant la 10 réponse fréquentielle  $H_{n,f}^3$  du filtre de débruitage par les composantes spectrales  $S_{n,f}$  du signal de parole (multiplieur 64). Un module 65 reconstruit alors le signal débruité dans le domaine temporel, en opérant la transformée de Fourier rapide inverse (TFRI) inverse des 15 échantillons de fréquence  $S_{n,f}^3$  délivrés par le multiplieur 64. Pour chaque trame, seuls les  $N/2=128$  premiers échantillons du signal produit par le module 65 sont délivrés comme signal débruité final  $s^3$ , après reconstruction par addition-recouvrement avec les  $N/2=128$  20 derniers échantillons de la trame précédente (module 66).

REVENDEICATIONS

1. Procédé de détection d'activité vocale dans un signal de parole numérique (s) traité par trames successives, dans lequel on soumet le signal de parole à un débruitage en tenant compte d'estimations du bruit compris dans le signal, mises à jour pour chaque trame d'une manière dépendante d'au moins un degré d'activité vocale ( $\gamma_{n,i}$ ) déterminé pour ladite trame, caractérisé en ce qu'on procède à un débruitage a priori du signal de parole de chaque trame sur la base d'estimations du bruit ( $\alpha'_{n-\tau_1,i} \cdot \hat{B}_{n-\tau_1,i}$ ) obtenues lors du traitement d'au moins une trame précédente, et on analyse les variations d'énergie du signal débruité a priori ( $\hat{E}_{p_{n,i}}$ ) pour détecter le degré d'activité vocale de ladite trame.
2. Procédé selon la revendication 1, dans lequel le degré d'activité vocale ( $\gamma_{n,i}$ ) est un paramètre non binaire.
3. Procédé selon la revendication 2, dans lequel le degré d'activité vocale ( $\gamma_{n,i}$ ) est une fonction, variant continûment entre 0 et 1.
4. Procédé selon l'une quelconque des revendications précédentes, dans lequel les estimations du bruit sont obtenues dans différentes bandes fréquentielles du signal, le débruitage a priori est effectué bande par bande, et il est déterminé un degré d'activité vocale ( $\gamma_{n,i}$ ) pour chaque bande.
5. Procédé selon l'une quelconque des revendications précédentes, dans lequel on obtient une estimation du bruit  $\hat{B}_{n,i}$  pour la trame n dans une bande de fréquences i sous la forme :

- 19 -

$$\hat{B}_{n,i} = \gamma_{n,i} \cdot \hat{B}_{n-1,i} + (1-\gamma_{n,i}) \cdot \tilde{B}_{n,i}$$

avec  $\tilde{B}_{n,i} = \lambda_B \cdot \hat{B}_{n-1,i} + (1-\lambda_B) \cdot S_{n,i}$

où  $\lambda_B$  est un facteur d'oubli compris entre 0 et 1,  $\gamma_{n,i}$  est le degré d'activité vocale déterminé pour la trame n dans la bande de fréquences i, et  $S_{n,i}$  est une moyenne de l'amplitude du spectre du signal de parole de la trame n sur la bande i.

6. Procédé selon la revendication 5, dans lequel le signal débruité a priori  $\hat{E}p_{n,i}$  relativement à une trame n et à une bande de fréquences i est de la forme :

$$\hat{E}p_{n,i} = \max \left\{ H p_{n,i} \cdot S_{n,i}, \beta p_i \cdot \hat{B}_{n-\tau_1,i} \right\}$$

où  $H p_{n,i} = \frac{S_{n,i} - \alpha'_{n-\tau_1,i} \cdot \hat{B}_{n-\tau_1,i}}{S_{n-\tau_2,i}}$ ,  $\tau_1$  est un entier au moins égal à 1,  $\tau_2$  est un entier au moins égal à 0,  $\alpha'_{n-\tau_1,i}$  est un coefficient de surestimation déterminé pour la trame  $n-\tau_1$  et la bande i, et  $\beta p_i$  est un coefficient positif.

7. Procédé selon l'une quelconque des revendications précédentes, dans lequel on calcule une estimation à long terme ( $\bar{E}_{n,i}$ ) de l'énergie du signal débruité a priori ( $\hat{E}p_{n,i}$ ), et on compare cette estimation à long terme à une estimation instantanée (ba) de cette énergie, calculée sur la trame en cours, pour obtenir le degré d'activité vocale ( $\gamma_{n,i}$ ) de ladite trame.

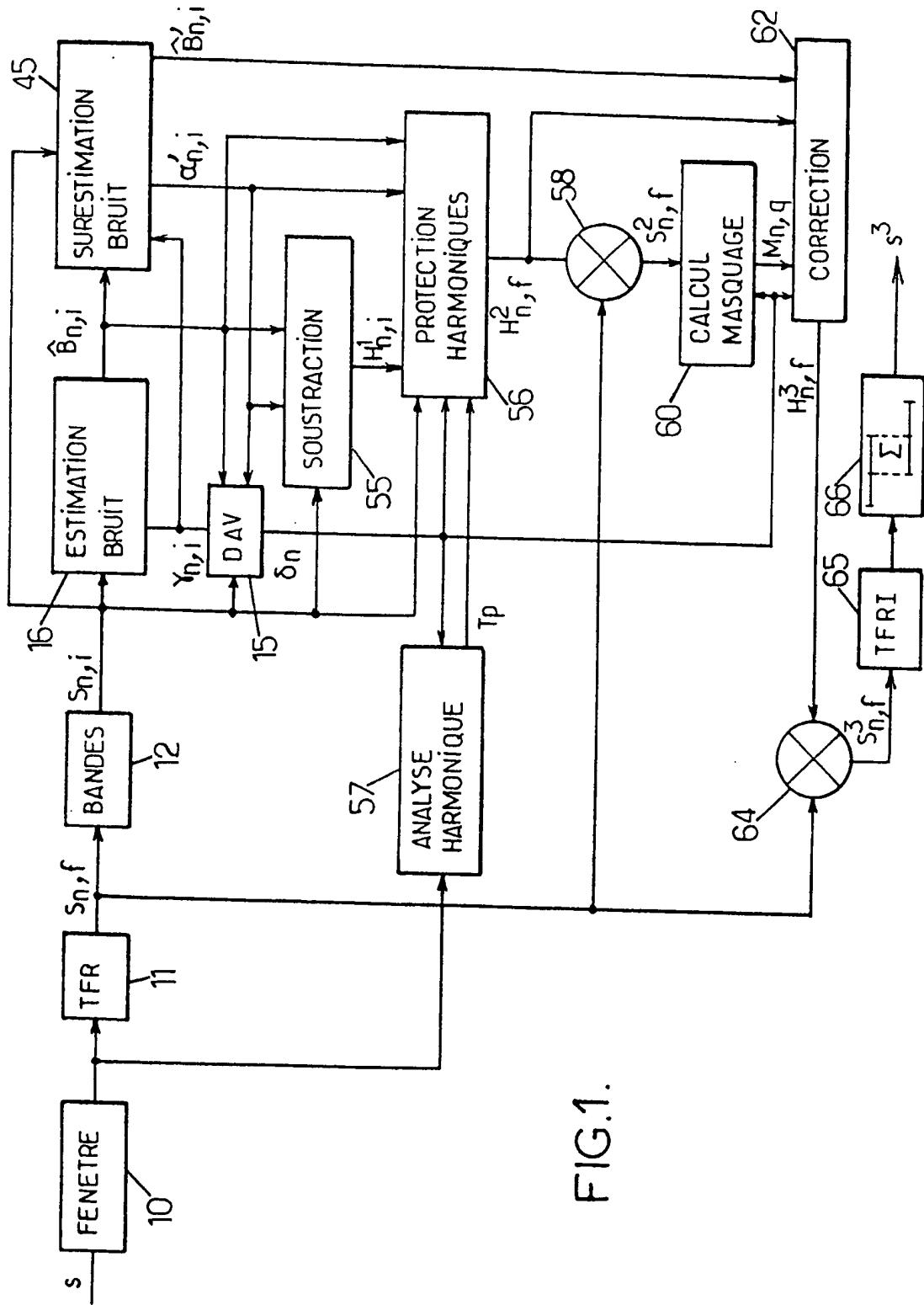


FIG.1.

2/5

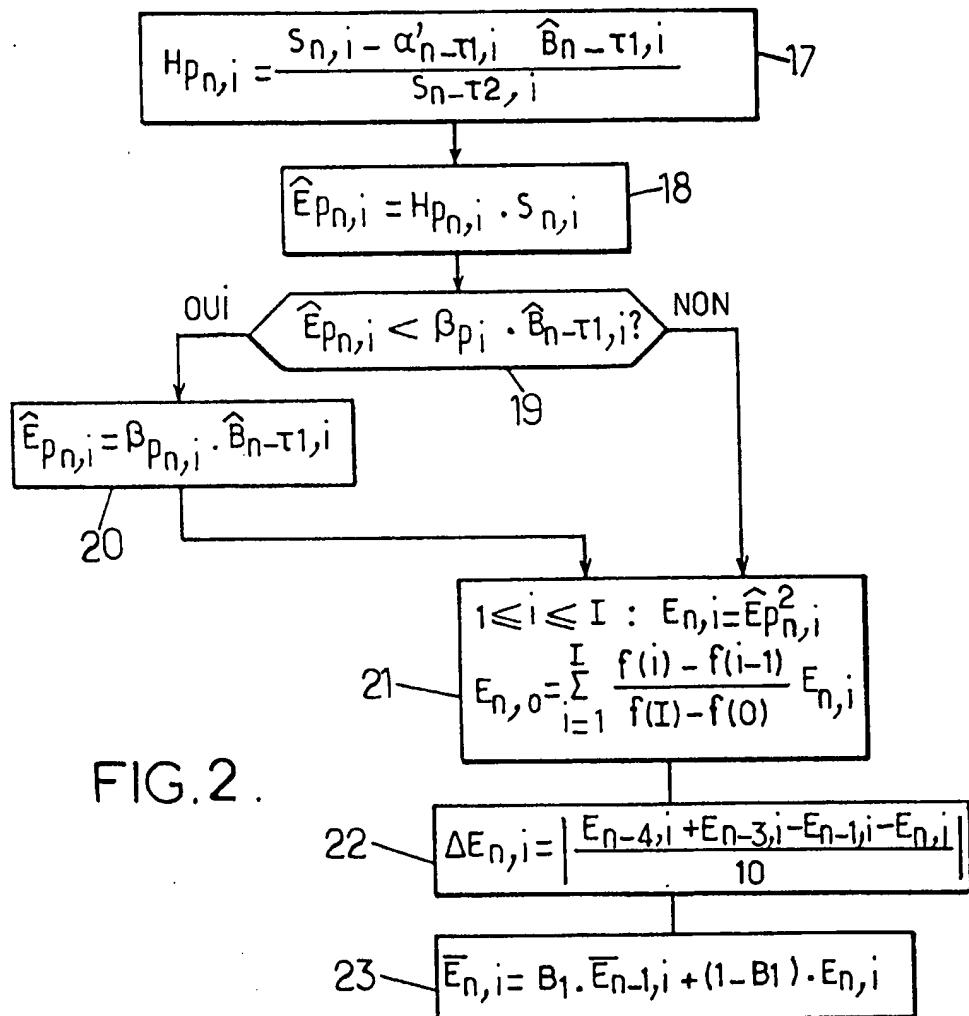
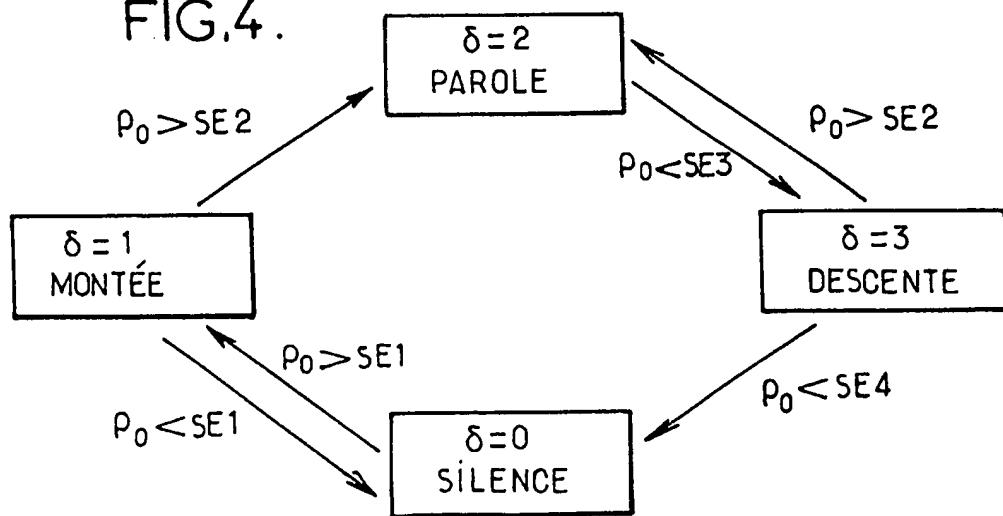
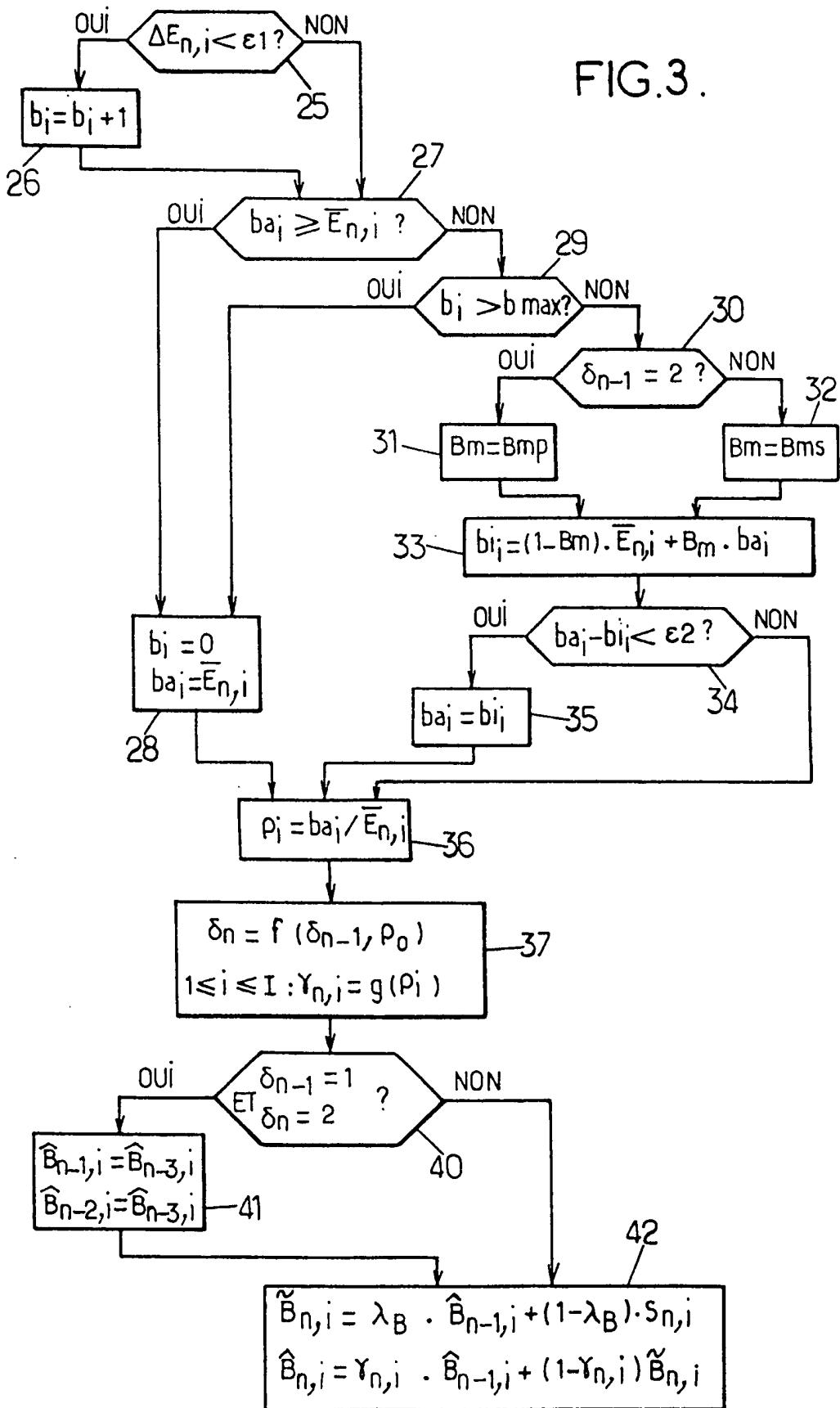


FIG.4.



3/5



4/5

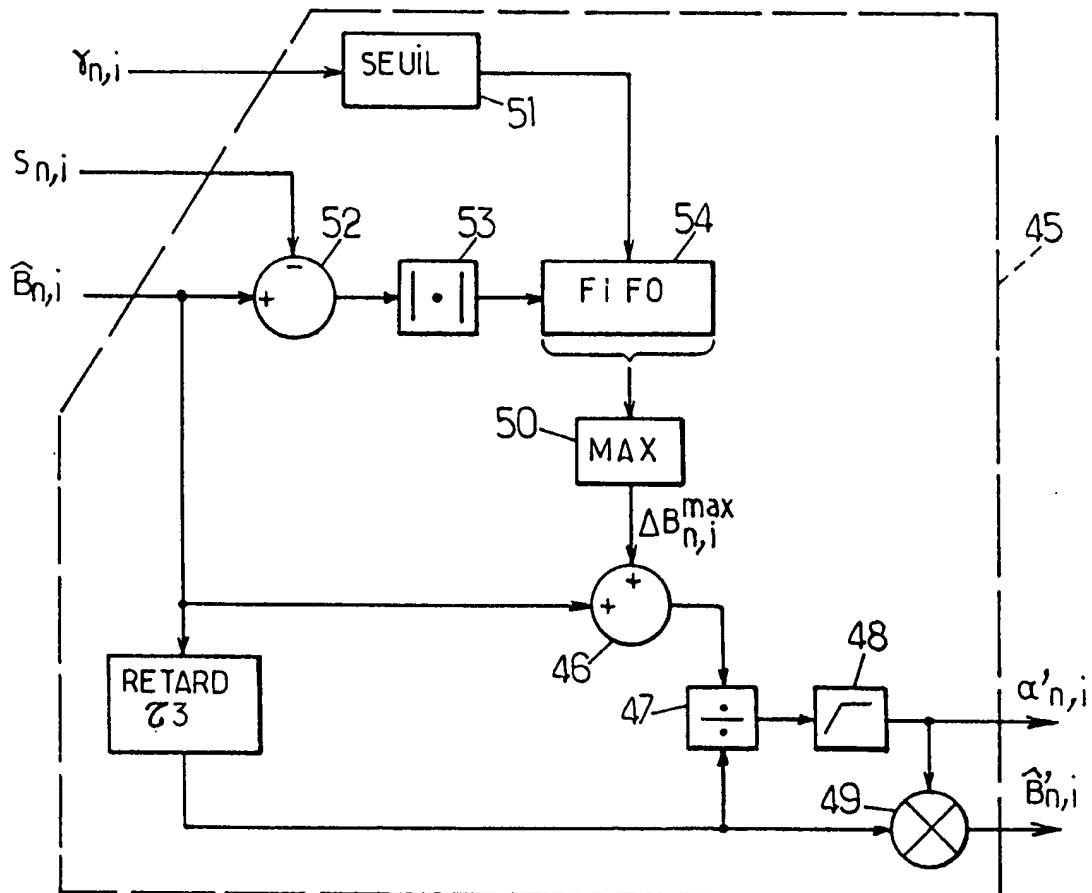
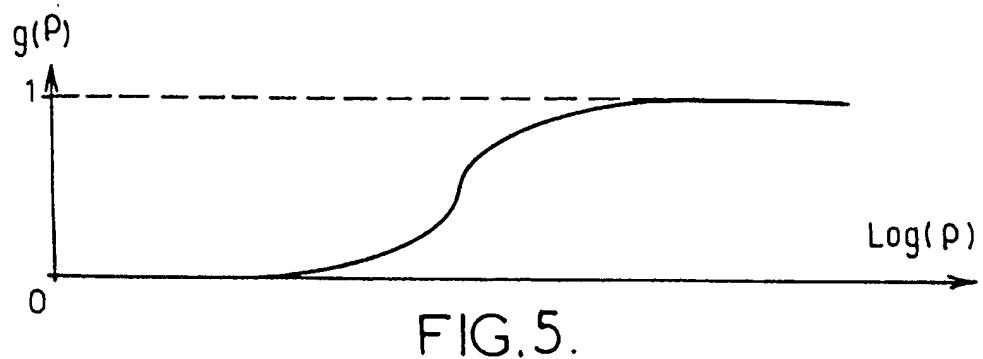


FIG. 6.

5/5

FIG. 7.

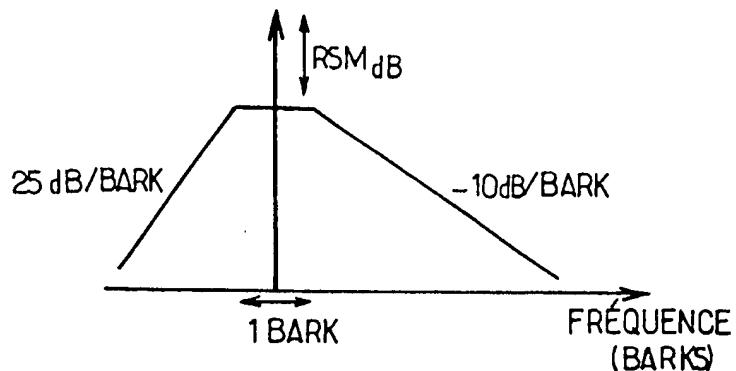
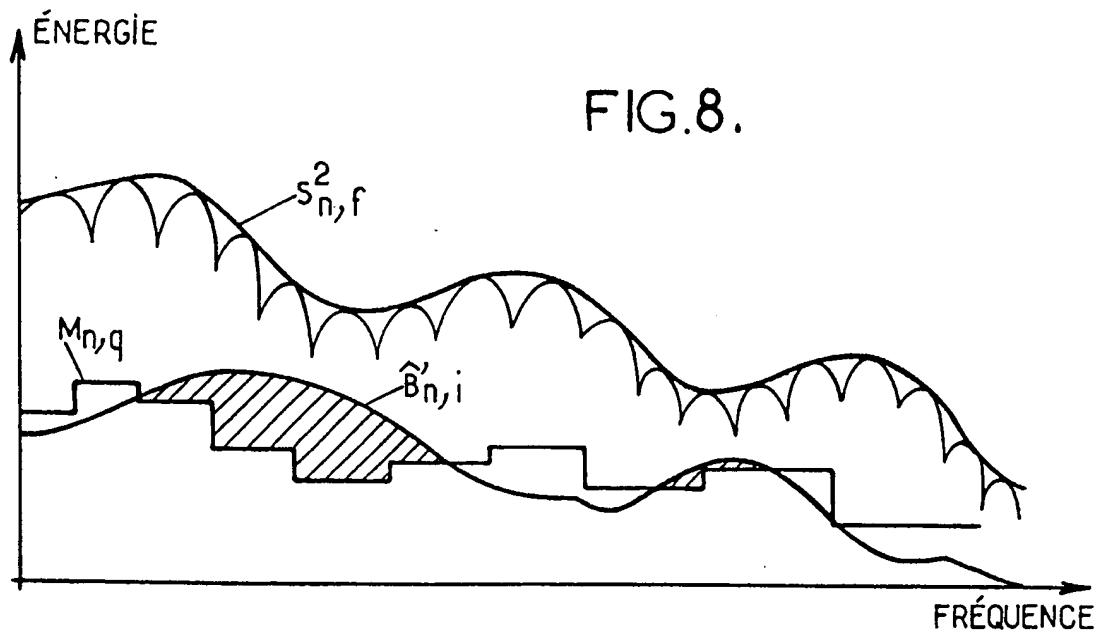


FIG. 8.



# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/FR 98/01979

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 6 G10L3/00

According to International Patent Classification(IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 659 622 A (ASHLEY) 19 August 1997 see column 3, line 12 - line 53 ---	1
A	PATENT ABSTRACTS OF JAPAN vol. 095, no. 006, 31 July 1995 & JP 07 074709 A (SONY), 17 March 1995 see abstract -& US 5 732 390 A (KATAYANAGI ET AL.) 24 March 1998 see column 2, line 17 - line 56 see column 6 - column 7, line 28 ---	1
A	DE 40 12 349 A (RICOH) 25 October 1990 see column 4 - column 5 -----	1

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

23 November 1998

Date of mailing of the international search report

30/11/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.  
Fax: (+31-70) 340-3016

Authorized officer

Lange, J

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/FR 98/01979

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
US 5659622	A 19-08-1997	AU 689403	B	26-03-1998
		AU 1758497	A	05-06-1997
		BR 9607249	A	30-12-1997
		CA 2203917	A	14-05-1997
		CN 1168204	A	17-12-1997
		DE 19681070	T	26-02-1998
		FI 972852	A	03-07-1997
		FR 2741217	A	16-05-1997
		GB 2313266	A	19-11-1997
		HU 9800843	A	28-07-1998
		SE 9701659	A	12-09-1997
		WO 9718647	A	22-05-1997
<hr/>				
DE 4012349	A 25-10-1990	JP 2278298	A	14-11-1990
		JP 3122699	A	24-05-1991
		US 5212764	A	18-05-1993
<hr/>				

# RAPPORT DE RECHERCHE INTERNATIONALE

De de Internationale No  
PCT/FR 98/01979

**A. CLASSEMENT DE L'OBJET DE LA DEMANDE**  
CIB 6 G10L3/00

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

**B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE**

Documentation minimale consultée (système de classification suivi des symboles de classement)

CIB 6 G10L

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)

**C. DOCUMENTS CONSIDERES COMME PERTINENTS**

Catégorie	Identification des documents cités, avec le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	US 5 659 622 A (ASHLEY) 19 août 1997 voir colonne 3, ligne 12 - ligne 53 ----	1
A	PATENT ABSTRACTS OF JAPAN vol. 095, no. 006, 31 juillet 1995 & JP 07 074709 A (SONY), 17 mars 1995 voir abrégé -& US 5 732 390 A (KATAYANAGI ET AL.) 24 mars 1998 voir colonne 2, ligne 17 - ligne 56 voir colonne 6 - colonne 7, ligne 28 ----	1
A	DE 40 12 349 A (RICOH) 25 octobre 1990 voir colonne 4 - colonne 5 -----	1



Voir la suite du cadre C pour la fin de la liste des documents



Les documents de familles de brevets sont indiqués en annexe

\* Catégories spéciales de documents cités:

- "A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- "E" document antérieur, mais publié à la date de dépôt international ou après cette date
- "L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- "O" document se référant à une divulgarion orale, à un usage, à une exposition ou tous autres moyens
- "P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée	Date d'expédition du présent rapport de recherche internationale
23 novembre 1998	30/11/1998
Nom et adresse postale de l'administration chargée de la recherche internationale Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl. Fax: (+31-70) 340-3016	Fonctionnaire autorisé  Lange, J

# RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
US 5659622 A	19-08-1997	AU 689403 B AU 1758497 A BR 9607249 A CA 2203917 A CN 1168204 A DE 19681070 T FI 972852 A FR 2741217 A GB 2313266 A HU 9800843 A SE 9701659 A WO 9718647 A	26-03-1998 05-06-1997 30-12-1997 14-05-1997 17-12-1997 26-02-1998 03-07-1997 16-05-1997 19-11-1997 28-07-1998 12-09-1997 22-05-1997
DE 4012349 A	25-10-1990	JP 2278298 A JP 3122699 A US 5212764 A	14-11-1990 24-05-1991 18-05-1993